

A COMPARATIVE AND HYBRID ENSEMBLE FRAMEWORK FOR CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING AND DEEP LEARNING TECHNIQUES

A Thesis Submitted
in Partial Fulfillment of the Requirements for the Degree of

MASTER OF TECHNOLOGY
IN
SOFTWARE ENGINEERING
BY
PRIYANKA
24/SWE/19

Under the Supervision of
Ms. Priya Singh
Assistant Professor, Department of Software Engineering
Delhi Technological University, Delhi, India



Department of Software Engineering
DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-110042, India

June, 2026



DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-110042

ACKNOWLEDGEMENT

I express my sincere gratitude to Ms. Priya Singh, Assistant Professor, Department of Software Engineering, Delhi Technological University, for her continuous guidance and mentorship throughout this dissertation. Her clarity in explaining the scope and objectives of this work, her willingness to address every difficulty as it arose, and her consistent encouragement were indispensable to its completion. Without her support and direction, this project would not have reached its present form.

I would also like to acknowledge the administrative and technical support extended by the Department of Software Engineering at Delhi Technological University. The computing infrastructure and access to research literature provided by the institution were essential resources for carrying out this work. I am also grateful to my fellow students in the department, whose conversations and observations helped me think through several problems more carefully than I would have managed alone.

Finally, I acknowledge the open-source communities and public dataset repositories whose freely available tools and data form the empirical foundation of this thesis.

Place: Delhi

PRIYANKA

Date:

24/SWE/19



DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I **Priyanka**, hereby certify that the work which is being presented in the thesis entitled "**A Comparative and Hybrid Ensemble Framework for Credit Card Fraud Detection Using Machine Learning and Deep Learning Techniques**" in partial fulfillment of the requirements for the award of the Degree of Master of Technology in Software Engineering, submitted in the Department of Software Engineering, Delhi Technological University is an authentic record of my own work carried out during the period from August 2024 to May 2026 under the supervision of Ms. Priya Singh.

The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other Institute.

Candidate's Signature

This is to certify that the candidate has incorporated all the corrections suggested by the examiners in the thesis and the statement made by the candidate is correct to the best of our knowledge.

Signature of Supervisor(s)

Signature of External Examiner



DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-110042

CERTIFICATE

This is to certify that the Project Dissertation titled “**A Comparative and Hybrid Ensemble Framework for Credit Card Fraud Detection Using Machine Learning and Deep Learning Techniques**” submitted by **PRIYANKA**, Roll No. **24/SWE/19**, Department of Software Engineering, Delhi Technological University, Delhi, in partial fulfillment of the requirement for the award of the degree of **Master of Technology** (Software Engineering), is a record of the project work carried out by the student under my supervision.

To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi

Ms. Priya Singh

Date:

Assistant Professor

Department of Software Engineering

ABSTRACT

Detecting credit card fraud is a critical issue within finance due to the problems associated with the extreme level of class imbalance, the continuously changing behaviour of fraudsters, and the need to make real-time decisions. This thesis aims to provide a contribution to improving machine learning's application for detecting credit card fraud. The first of these papers provides a structured comparative analysis of supervised, unsupervised, deep learning, ensemble and hybrid algorithm-based methods for detecting credit card fraud. For the purpose of comparison, algorithms such as logistic regression, random forest, support vector machine, gradient boosting, autoencoder, and isolation forest were compared according to different criteria such as the characteristics of datasets employed, approaches to class imbalance through SMOTE and ADASYN, and multiple performance metrics (i.e. Precision, Recall, F1-Score, and AUC). The second contribution presents a new threshold-tuned hybrid ensemble methodology based on Random Forest, LightGBM and Autoencoder utilizing no synthetic oversampling technique. This study used an exhaustive search technique for optimising the weights assigned to each classifier and the decision threshold in order to maximise the F1-Score on the original data with the class imbalances. The framework was tested using the datasets for credit card fraud detection available through the Kaggle. The results of this method were impressive, yielding overall performance metrics: 99.96% accuracy, 0.9318 precision, 0.8367 recall, 0.8817 F1 score and 0.9756 ROC-AUC with this method of approach outperforming all individual classifiers.

LIST OF PUBLICATIONS

Published/Communicated

INTERNATIONAL CONFERENCES

1. Priyanka & Priya Singh (2025). A Hybrid Learning Strategy for Credit Card Fraud Detection Using Calibrated Thresholds. [**Scopus Indexed**][**Accepted and Presented**]
2. Priyanka & Priya Singh (2026). A Comparative Analysis of Machine Learning and Deep Learning Approaches for Credit Card Fraud Detection under Class Imbalance: A Review. [**Scopus Indexed**][**Accepted and Presented**]

TABLE OF CONTENTS

Acknowledgement	i
Candidate's Declaration	ii
Certificate	iii
Abstract	iv
List of Publications	v
Table of Contents	vi
List of Tables	ix
List of Figures	x
List of Symbols and Abbreviations	xi
1 INTRODUCTION	1
1.1 Background and Motivation	1
1.2 Problem Statement	2
1.3 Research Objectives	3
1.4 Scope of the Thesis	4
1.5 Key Contributions	4
1.6 Thesis Organization	5
1.7 Chapter Summary	6
2 LITERATURE REVIEW	8
2.1 Anomaly Detection in Credit Card Fraud	8
2.2 Machine Learning Paradigms in Fraud Detection	9
2.2.1 Supervised Learning Approaches	9
2.2.2 Unsupervised Learning Approaches	9

2.2.3	Deep Learning Techniques	10
2.2.4	Ensemble and Hybrid Models	10
2.3	Big Data Challenges in Credit Card Fraud Detection	10
2.4	Credit Card Fraud Detection Process and Modeling Framework	11
2.4.1	Data Preparation and Feature Engineering	11
2.4.2	Class Imbalance Handling	11
2.4.3	Model Training Approaches	12
2.4.4	Evaluation Metrics	12
2.5	Existing Techniques and Comparative Analysis	12
2.5.1	Synthesis and Discussion	14
3	EXPERIMENTAL DESIGN	15
3.1	Proposed Methodology	15
3.1.1	Overall Framework Architecture	15
3.1.2	Data Preprocessing and Splitting Strategy	16
3.1.3	Selected Baseline Models	17
3.2	Dataset Description	20
3.2.1	Dataset Overview	20
3.2.2	Feature Description	21
3.2.3	Data Splitting Summary	22
3.3	Hybrid Ensemble Construction	22
3.4	Threshold Tuning and Optimization	23
3.5	Performance Measures	23
3.6	Parameter Settings	25
3.6.1	Random Forest Parameters	26
3.6.2	Autoencoder Parameters	26
3.6.3	LightGBM Parameters	27
3.6.4	Hybrid Ensemble and Optimization Parameters	27
4	EXPERIMENTAL RESULT AND ANALYSIS	29
4.1	Results	29
4.1.1	Precision-Recall and ROC Curve Analysis	30
4.1.2	Confusion Matrix	31
4.2	Analysis	32
4.2.1	Interpretation of the Optimal Weight Vector	32
4.2.2	Interpretation of the Optimal Decision Threshold	33
4.2.3	Computational Complexity	33

5	LIMITATIONS	35
6	CONCLUSION AND FUTURE SCOPE	37
6.1	Conclusion	37
6.2	Future Work	38
	Bibliography	40
	Declaration	44
	Publication Details and Proofs	46
	Plagiarism Verification	53
	Plagiarism Report	54
	Author's Biography	58

LIST OF TABLES

2.1	Theoretical comparative study of research papers	13
2.2	Big Data and Dataset Comparative Study	13
2.3	Performance Metrics Comparison	14
3.1	Summary Statistics of the Kaggle Credit Card Fraud Detection Dataset	21
3.2	Stratified Data Splitting Summary (fraud ratio $\approx 0.172\%$ preserved in each split)	22
3.3	Random Forest Hyperparameter Settings	26
3.4	Autoencoder Architecture and Training Hyperparameters	26
3.5	LightGBM Hyperparameter Settings	27
3.6	Hybrid Ensemble and Optimization Parameters	27
4.1	Performance Metrics for the Models on the Test Set	29
4.2	Confusion Matrix of the Optimized Hybrid Ensemble on the Test Set .	31

LIST OF FIGURES

2.1	Workflow of the Credit Card Fraud Detection System	12
3.1	Pipeline Overview of the Proposed Hybrid Ensemble Framework . . .	16
3.2	Architecture of Random Forest	18
3.3	Architecture of Autoencoder	19
3.4	Architecture of LightGBM	20
4.1	PR Plot for Hybrid Ensemble Model	30
4.2	ROC Plot for Hybrid Ensemble Model	31
4.3	Computational Complexity	34

LIST OF SYMBOLS AND ABBREVIATIONS

Abbreviation	Full Form
AE	Autoencoder
AI	Artificial Intelligence
ADASYN	Adaptive Synthetic Sampling
ANN	Artificial Neural Network
AP	Average Precision
AUC	Area Under Curve
CNN	Convolutional Neural Network
DL	Deep Learning
FN	False Negative
FP	False Positive
KNN	K-Nearest Neighbors
LGBM	Light Gradient Boosting Machine
ML	Machine Learning
MSE	Mean Squared Error
PCA	Principal Component Analysis
PR	Precision-Recall
RF	Random Forest
ROC	Receiver Operating Characteristic
SMOTE	Synthetic Minority Oversampling Technique
TN	True Negative
TP	True Positive

Chapter 1

INTRODUCTION

1.1 Background and Motivation

During the last twenty years, there has been a huge change in the world of finances due to fast digitization of the financial payment system [18, 19]. Thanks to the development of numerous electronic commerce websites, mobile banking applications, and even digital wallets, credit card payments have now become a critical component of the contemporary business environment, responsible for billions of financial transactions conducted every day around the world. Of course, these achievements have reduced costs and increased the convenience of financial transactions; but at the same time, they have created some additional threats that are used by many criminal actors. How surprising is the price paid for such fraud? According to some estimates, total losses resulting from these activities reach \$32.34 billion and are still growing because of electronic transactions [23]. What is more, besides economic consequences, the reputation damage related to fraudulent activities may include damage done to financial institutions and a rather costly process of investigation needed to solve the fraud issue. Different kinds of fraudulent schemes exist, ranging from lost/stolen cards, CNP (card-not-present) fraud, large-scale hacking attacks, and phishing to identity frauds [18, 19]. Classical fraud detection tools have been developed using rule-based mechanisms whereby the rules involve thresholds and heuristic approaches to flag suspicious behaviors. Although these systems offer some level of protection against fraud, they also exhibit certain weaknesses when it comes to advanced attacks [20, 24]. The reactive property of rule-based systems means that the system can only detect the attack pattern if the behavior is already defined beforehand. Another limitation of rule-based fraud detection systems is the inability to generalize since the attackers are clever to design a method that goes beyond the rule definition. Furthermore, rule updates are manual and therefore not suitable for real-time detection [18, 24]. This issue led to developments in the area of financial technologies and machine learning to develop new

ways of detecting suspicious behaviors through data-driven approaches [11, 19]. The idea here is quite different from the one discussed above. In other words, we refer to machine learning techniques that utilize historical data of transactions to learn discriminatory patterns [6, 7, 13]. With supervised learning models, patterns of fraud can be inferred by capturing statistical regularities in a large dataset [8, 9, 10]. Unsupervised and anomaly-based approaches do not need any examples of fraud; they can detect deviations from normal behavior, and are better suited for detecting novel and zero-day fraud attempts [1, 2, 3, 4]. This thesis lies in the cross-section of the two paradigms, and it not only provides a thorough overview of current research but also introduces a new high-performance hybrid detection framework [15, 25, 27].

1.2 Problem Statement

Although there is a significant amount of research work in credit card fraud detection using machine learning, there are several crucial issues that are not well addressed [19]. There are three dimensions of interrelated problems that can be used to break down the central problem studied in this thesis.

Class Imbalance: Credit card transaction data sets are very imbalanced in the real world, with fraud transactions making up less than 0.2% of the total transactions [17]. The “European credit card dataset” on Kaggle is the most popular evaluation resource in the literature with 492 fraudulent transactions among 284,807 transactions (or a fraud ratio of 0.172%) [11, 15, 18, 22]. This extreme class imbalance makes the normal classifiers train extremely biased on the majority class, and so the trained classifiers have a very high accuracy, but are not able to detect the majority of the real fraud cases. This imbalance, however, without any distortion of the underlying data distribution by the system, is a fundamental open challenge.

Model Generalization and Fragmentation: The existing literature is highly fractured and individual research studies assess specific algorithms in different experimental setups, with different datasets and evaluation methods [6, 11, 17, 19]. This fragmentation makes it difficult to compare the performance of the algorithms across different studies and hinders clear comparison of the relative strengths and weaknesses of the different algorithmic families: supervised classifiers, unsupervised anomaly detectors, deep learning architectures and hybrid ensemble methods. A thorough, systematic comparison is therefore required to integrate state of knowledge and target the most promising research directions.

Threshold Optimization and Practical Deployment: For most of the existing hybrid and ensemble models, the models are based on synthetic oversampling methods

like SMOTE [5, 12, 22] (which may add patterns to the data and thus decrease real-world generalizability) or use default decision thresholds that may not necessarily be optimal for imbalanced binary classification problems [25, 27]. Even with the use of synthetic augmentation, optimizing the weights of the model components and decision thresholds to maximize F1 score on the original imbalanced dataset is not thoroughly explored and is a key difference between research models and practically deployed fraud detection models.

1.3 Research Objectives

The main research goals of this thesis are:

- To perform a thorough and systematic analysis of ML and DL methods for detecting credit card fraud, including supervised, unsupervised, ensemble and hybrid techniques and focus on techniques for handling class imbalance and evaluation frameworks.
- To recognize major theoretical and empirical drawbacks of current methods and analyze them in structured way in comparison with algorithmic design, properties of the datasets, big data challenges, and performance measures reported in recent publications.
- To develop, deploy, and thoroughly test a new hybrid ensemble model based on thresholds which combines complementary supervised classifiers (Random Forest and LightGBM) with an unsupervised deep learning anomaly detector (Autoencoder) for credit card fraud detection.
- To show that grid search for optimization of model component weights and optimization of decision thresholds on a held-out validation set can yield better fraud detection performance on highly imbalanced data than can oversampling synthetic data.
- To compare the proposed hybrid model to constituent models as well as to the related previous works, an extensive performance evaluation is performed using the following performance metrics, which are suitable for imbalanced classification: Precision, Recall, F1-score, ROC-AUC, and Confusion Matrix Analysis.
- To examine computational characteristics of the proposed framework in terms of training time and memory usage, and determine the practical feasibility of the framework to be deployed in the real financial transaction environment.

1.4 Scope of the Thesis

This thesis is about the binary credit card fraud detection problem, which is the classification of individual financial transactions as fraud or not fraud. The experimental studies are performed on the publicly available Kaggle European credit card transaction data set [11, 15, 18, 22], which is a commonly used test set for the Research Community in Fraud Detection. This is a strict and repeatable assessment framework, but the thesis notes that it is necessary to generalize to other regional data sets, transaction types and environments where data is streaming in real-time. The Literature Review includes supervised learning algorithms like Logistic Regression, Decision Tree, Random Forest, SVM, K-Nearest Neighbor algorithm, and Gradient Boosting algorithms [6, 7, 8, 10]; unsupervised and outlier detection algorithms like Isolation Forest, Support Vector Data Description, K-means clustering, and Autoencoders [1, 2, 3, 4]; DL algorithms like ANN and CNN [5, 14, 15]; and other hybrid and ensemble learning algorithms [16, 25, 27]. The review is based on research articles appearing during the last three years (2022–2025), which represent the state of the art. This choice of not using synthetic oversampling is a fundamental methodological difference with most previous work [5, 12, 22], as it allows the proposed experimental framework to be used with models trained and evaluated based on the original distribution of imbalanced data.

1.5 Key Contributions

The following are original contributions to Machine learning-based financial fraud detection made in this thesis:

- **Systematized Comparative Survey:** A detailed survey of twenty-two recent research papers addressing credit card fraud detection is given. Unlike the existing survey papers that make an aggregated presentation of the results without any methodological rigor, this survey presents the results across three parallel dimensions: (i) theoretical comparison of algorithmic approach, contributions and limitations; (ii) dataset and big data challenge analysis with class imbalance ratios, feature characteristics and real-time detection capabilities; and (iii) quantitative performance comparisons in terms of Accuracy, Recall, F1-score and AUC. This multi-dimensional synthesis enables a better understanding of the relative state of various algorithmic families and leaves and explicitly points out the research gaps which inspire the experimentation.
- **Novel Threshold-Tuned Hybrid Ensemble Framework:** A practical and high-

performance hybrid fraud detection model based on a combination of three complementary models: a Random Forest classifier, a LightGBM classifier, and a deep Autoencoder-based anomaly detector, is proposed. The individual models are ranked and then a hybrid model score is formed through the weights giving the combined output. A full grid search is then carried out over both weights of the model components and decision threshold values, and the maximum F1 score is found on a held-out validation set. Most importantly, this whole framework is based on the original imbalanced data with no synthetic oversampling, which results in a more realistic and generalizable model than SMOTE dependent models.

- **Empirical Demonstration of Synergistic Model Integration:** This thesis shows that the threshold-tuned hybrid ensemble obtains a higher F1-score (0.8817), Precision (0.9318), Recall (0.8367), and ROC-AUC (0.9756) than each individual constituent model, and even than previous SMOTE-based deep learning methods reported in the literature. The best weight configuration (WRF = 0.5, WAE = 0.4, WLGBM = 0.1) empirically shows the strong complementary role of the unsupervised Autoencoder to identify anomalous transactions that are not caught by the supervised classifiers, providing evidence for the synergistic potential of supervised and unsupervised learning in a single framework for scoring transactions.
- **Practical Deployment Analysis:** A computational complexity analysis of training time and memory usage for all model components is provided, showing that the hybrid ensemble introduces very little overhead over the model scores that are pre-computed. Random Forest has the highest training time, LightGBM has the highest memory usage, but the Autoencoder is able to make the inference faster, which is essential in streaming transaction environments. The data obtained from the research corresponds to the practical application of the framework developed in the previous sections to the financial sector.

1.6 Thesis Organization

This thesis is divided into the following:

Chapter 2 — Literature Review: The literature reviewed in this chapter summarizes current ML and DL research to detect credit card fraud. Algorithms are categorized by learning paradigm and the review combines the results of comparative studies of algorithm design, data characteristics, imbalance handling approaches, and evaluation

metrics. The final section of the chapter states the limitations associated with the research that formed the basis for developing the methodology proposed.

Chapter 3 — Experimental Design: The complete design and implementation of the threshold-tuned hybrid ensemble framework is detailed, including the description and pre processing of the datasets, the splitting strategy, individual model architectures and training methods, and the ensemble and the process of optimizing the weights and thresholds.

Chapter 4 — Experimental Result and Analysis: Results from the held-out test set are presented and analyzed in detail. The individual model baselines are compared to the hybrid ensemble. In addition to the measurements of computational complexity, ablation studies, and threshold sensitivity analysis, confusion matrix analysis, PR curve evaluation, and ROC-AUC assessment are also provided.

Chapter 5 — Limitation: The limitations of hybrid fraud detection framework are outlined in this chapter. The limitations include reliance on a single benchmark data set, no evaluation of real-time deployment, poor generalization capability for different financial ecosystems, and issues in computational complexity for large-scale transaction streams. In addition, the limitations are also covered with respect to changing fraud patterns and the lack of explainable techniques for deep learning based fraud detection models.

Chapter 6 — Conclusion and Future Work: In conclusion, this thesis offered a synthesis of major contributions and findings, outlining limitations, along with recommendations for future research directions (e.g., real-time system deployment, validation with diverse datasets, implementation of adaptive/learning development frameworks, and development of explainability and regulatory compliance frameworks).

1.7 Chapter Summary

The motivation for the research presented in this thesis has been set out in this chapter, which has been influenced by the increasing financial and operational losses due to fraud committed with credit cards and the lack of effectiveness of the current rule-based and isolated machine learning solutions. These three issues — 'highly imbalanced classes', 'scattered knowledge of the different methods', and 'differences between research models and systems that can be used in practice' — have been well expressed as the motivating issues that should be tackled by the work. The chapter has clearly stated the research objectives, clearly set forth the scope and limits of in-

vestigation, summarized the major original contributions to the field, and outlined the roadmap of the thesis. The following chapters elaborate the theoretical, empirical and analytical aspects of this study in greater detail.

Chapter 2

LITERATURE REVIEW

Due to the growth of machine learning tools and data analytics, credit card fraud detection is evolving. To understand how effective detection systems were created, we must investigate both the theoretical underpinnings of anomaly detection as well as the ML approaches applied to fraud detection in financial transactions. These research contributions can help explore all theoretical aspects of these approaches by utilizing supervised, unsupervised, deep learning, and hybrid methods.

2.1 Anomaly Detection in Credit Card Fraud

Anomaly detection is act of identifying abnormal incidents or events that are out of the normal data trend. Fraudulent activities in credit card transactions are viewed as an abnormality relative to the normal spending patterns. The fact that percentage of fraudulent transactions lies in a limited spread of total transactions implies that anomaly detection methods are significant in the process of detecting any form of suspicious behavior. The methods used in some of the studies are unsupervised or semi-supervised anomaly detection. As an example, the Support Vector Data Description hybrid with fuzzy logic [1] is used to describe normal transaction patterns and identify fraudulent activity. Equally, the hybrid Isolation Forest method [20] detects the abnormal pattern of transactions by isolating abnormalities. Autoencoder based anomaly detection, e.g., the intelligent learning scheme [19] and the MDIASE Autoencoder [21] use normal transaction patterns, which are patternized and then transactions with high reconstruction error are considered anomalies. Overall, credit card transaction data anomalies can be divided into three major categories:

1. Point anomalies, these are where a single transaction has a large deviation with respect to normal behaviour.

2. The contextual anomalies, where a transaction occurs abnormally in a given context in terms of time or behaviour.
3. Group anomalies, in which a cluster of transactions involving a group of similar transactions has a sign of fraud.

Autoencoder and hybrid models [12, 19] particularly demonstrate effectiveness in capturing complex nonlinear patterns, enabling improved anomaly detection compared to traditional statistical methods.

2.2 Machine Learning Paradigms in Fraud Detection

Techniques in ML theories applied in credit card fraud are broadly divided into supervised, unsupervised, deep learning, ensemble, and hybrid.

2.2.1 Supervised Learning Approaches

Usually, credit card fraud detection tends to use supervised learning, which is by far the most widely used in research studies. Common methods include logistic regression, decision trees, random forests, SVM, KNN, and gradient boosting [3, 6, 11, 13]. Random Forest is highly accurate and precise in various studies [15, 18]. Since Logistic Regression is often reported to be simple and stable, especially when used with high-dimensional PCA-transformed features, this is a common practice in statistics and machine learning studies [11, 14]. Gradient Boosting and XGBoost models are the advanced predictive models, which minimize the errors in classification through iteration [14].

2.2.2 Unsupervised Learning Approaches

Unsupervised learning methods are specifically applicable to cases where there is limited labelled data or cases where the fraud patterns are continually changing. Methods and algorithms in anomaly detection, like Isolation Forest [20] and SVDD [1], do not use fixed labels of the classes to help cluster them. Such methods can detect new fraud patterns that have never been detected before. Nevertheless, unsupervised models alone can have a high rate of false positive results, unless tuned. Consequently, a number of studies use an unsupervised detector of anomalies with supervised classifiers to boost robustness.

2.2.3 Deep Learning Techniques

Neural networks such as ANN and CNN have been considered to identify nonlinear correlations of transaction features. The ANN-based framework [10] and the comparative deep learning analysis [14] prove to be competitive in terms of the AUC and F1-score. Moreover, [19, 21] exhibits a better ability to deal with highly imbalanced data through autoencoders because of their ability to model the internal structure of legitimate transactions.

Even though they work well, deep learning models cannot be used in real time due to large computational requirements and a need for hyperparameter fine-tuning.

2.2.4 Ensemble and Hybrid Models

Hybrid and ensemble models are learning algorithms that integrate several learning algorithms to enhance the performance of classification and minimize false alarms. The hybrid machine learning model [5] incorporates the use of more than one classifier in an effort to obtain a higher fraud detection rate. On the same note, the ensemble modelling techniques [17] exhibit higher stability through the aggregation of the predictions of various base learners.

An example of using a hybrid of the anomaly detection and classification models for improvement of performance of fraud detection models is hybrid Isolation Forest-K-Means-KNN model [20]. These strategies are used to trade off between detection and generalisation.

2.3 Big Data Challenges in Credit Card Fraud Detection

Big data challenges play a major role in development of an efficient system of fraud detection. The four key challenges that have been identified in the studies are:

1. **Class Imbalance (Value Challenge):** The severe imbalance between the legitimate and fraudulent transactions is one of the most important problems. Some studies specifically discuss this problem where SMOTE and ADASYN resampling methods are used to create artificial minority samples [12, 15, 22]. In absence of such methods, the models will be biased towards the majority group.
2. **Volume Challenge:** A large number of studies use the Kaggle European dataset of 284,807 transactions having PCA-transformed features [11, 15, 18]. To pro-

cess such large volumes, efficient preprocessing and feature selection methods are needed.

3. Velocity Challenge (Real-Time Detection): There is under research in real time fraud detection. Despite fact that some of the works [16] have banking-based systems, the majority of the experimental assessment is conducted in batch-processing contexts.
4. The methods of fraudsters are ever-changing; they need scalable and adaptive models. The hybrid and anomaly-based systems [1, 19, 21] are designed to solve this drawback of adaptability.

2.4 Credit Card Fraud Detection Process and Modeling Framework

The chosen articles mostly adhere to a general structure of fraud detection pipeline, which includes data preprocessing, imbalance management, model training, and performance evaluation. Although the algorithmic choices in different studies are different, the modeling frameworks have a consistent concept.

2.4.1 Data Preparation and Feature Engineering

The majority of research makes use of the European Kaggle credit card dataset that have 284,807 transactions with PCA-transformed variables [11, 15, 18, 22]. Since the convergence of the model is to be stable, feature scaling and normalization are used. Redundancy and complexity of computation dimensionality reduction are specifically addressed through PCA [4, 15].

2.4.2 Class Imbalance Handling

A large imbalance between the classes is a significant problem with fraud detection since fraudulent transactions are usually less than 0.2 percent of the total transactions. Several studies attempt to solve this problem by applying SMOTE and ADASYN to create synthetic minority samples that can better recall and achieve higher F1-score [12, 15, 22]. Instead, an imbalance in anomaly detection models like Autoencoders [19, 21] and those that use SVDD-based models can be implicitly tackled by modelling normal transaction behaviour.

2.4.3 Model Training Approaches

In the training phase, machine learning algorithms are implemented on transaction data that has been labelled. Managed classifiers such as Random Forest, Logistic Regression, Decision Trees, SVM and Gradient Boosting are actively used [11, 14, 15]. The hybrid and anomaly-based methods[1,20] and Autoencoder models, especially Deep learning models [19, 21], are proposed, which involve the mechanisms of classification and anomaly detection.

2.4.4 Evaluation Metrics

Performance evaluation goes further because of the imbalance. The majority of the studies indicate Precision, Recall, F1-score, and AUC [15, 19, 21]. Recall and F1-score are highlighted as more valid measures of fraud detection ability.

2.5 Existing Techniques and Comparative Analysis

In the research papers, the workflow of detection of the fraud can be shown in Figure 2.1. This systematic design is continuously seen throughout supervised methods [4, 6], anomaly-based methods [1, 20] and deep learning research [19, 21].

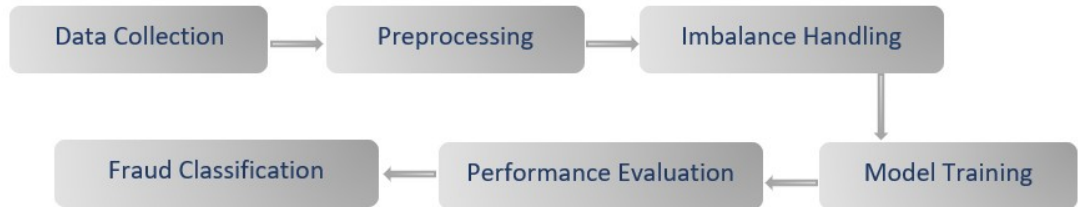


Figure 2.1: Workflow of the Credit Card Fraud Detection System

The results of the methodological comparison of the chosen works are summarised in Table 2.1, where the theoretical positions and the main approaches of research are provided. Dataset characteristics and big data considerations are summarised in Table 2.2. Table 2.3 summarises performance measures in the studies. Random Forest and boosting-based methods possess a high level of consistency as indicated in Table 2.3. Conversely, SMOTE-based methods raise recall and F1-score, especially in datasets that are extremely imbalanced. In addition, autoencoder and hybrid models show high AUC results in case of imbalance.

A data-driven comparison shows that the accuracy of ML models is between 85% to 93%. On the other hand, the accuracy of DL models is between 94% to 98%.

Moreover, deep learning models perform better when the data is imbalanced, whereas machine learning models are more efficient in real time applications. This indicates that higher accuracy is achieved at the cost of increased computational complexity.

Table 2.1: Theoretical comparative study of research papers

Ref	Primary Approach	Contribution	Limitation
[1]	SVDD + Fuzzy Logic	Boundary-based anomaly modeling without heavy supervision	Does not explicitly address severe class imbalance; limited benchmark comparison
[4]	Random Forest + PCA	High accuracy using ensemble learning with dimensionality reduction	Dependent on PCA-transformed dataset; limited generalization
[6]	Comparative Supervised ML	Demonstrates superiority of RF over single classifiers	Risk of overfitting; limited imbalance strategy
[10]	ANN-based Model	Captures nonlinear feature interactions	High computational cost; hyperparameter sensitivity
[12]	SMOTE + Deep Learning	Improved recall under imbalanced dataset	Synthetic sampling may introduce noise; overfitting risk
[15]	SMOTE + XGBoost	Achieved high AUC and balanced F1-score	Heavy dependence on Kaggle dataset; no real-time validation
[17]	Ensemble Modeling	Improved stability via multi-classifier aggregation	Increased model complexity; limited deployment analysis
[19]	Autoencoder-based Detection	Strong anomaly detection using reconstruction error	Threshold selection complexity; requires tuning
[20]	IF + KMeans + KNN	Hybrid anomaly-classifier integration	Computational overhead; scalability concerns
[21]	MDIASE Autoencoder	Enhanced latent representation learning	Model training complexity; dataset dependency
[22]	SMOTE-based Supervised ML	Improved recall and F1-score using resampling	Synthetic data bias; no adaptive learning

Table 2.2: Big Data and Dataset Comparative Study

Ref	Dataset	Features	Fraud Ratio	Imbalance Handling	Real-Time
[11]	Kaggle EU	30	$\sim 0.17\%$	No	No
[12]	Kaggle EU	30	$\sim 0.17\%$	SMOTE	No
[15]	Kaggle EU	30	$\sim 0.17\%$	SMOTE, ADASYN	No
[16]	Banking Dataset	Custom	Not specified	Limited	Partial
[19]	Kaggle EU	30	$\sim 0.17\%$	Implicit (AE)	No
[21]	Kaggle EU	30	$\sim 0.17\%$	Implicit (AE)	No
[22]	Kaggle EU	30	$\sim 0.17\%$	SMOTE	No

Table 2.3: Performance Metrics Comparison

Ref	Model	Accuracy	Recall	F1-score	AUC
[4]	Decision Tree	99.78%	–	–	–
[12]	SMOTE + Deep Learning	99.93%	90.2%	78.19%	–
[15]	Random Forest + ADASYN	–	–	86.77%	0.99
[19]	Autoencoder (ILSDFD)	98.96%	93.83%	–	–
[20]	Hybrid IF-KNN	99.68%	100%	99.68%	–
[21]	MDIASE-AE	–	–	–	0.999
[22]	SMOTE + Random Forest	99.95%	82.65%	84.82%	–

2.5.1 Synthesis and Discussion

As can be seen in the comparative analysis provided in Table 2.1, Table 2.2 and Table 2.3, the use of supervised ensemble techniques, in particular, the Random Forest and the boosting-based models, still prevails [4, 6]. The imbalances are also addressed with SMOTE, which significantly improves recall and F1-score [12, 15, 22]. Hybrid and anomaly detection approaches are more flexible to irregular patterns of fraud usually under extreme imbalance circumstances [1, 19, 20, 21]. However, excessive reliance on the Kaggle benchmark data is a limiting factor to the generalizability [11, 15, 18]. Other areas that are yet to be explored well are real time deployment and adaptive learning. Nevertheless, some obstacles are connected to the process including the issue of class Asymmetric distribution, the need to make the process real-time, and the lack of available standard data to compare.

Chapter 3

EXPERIMENTAL DESIGN

3.1 Proposed Methodology

This chapter gives a complete description of the hybrid ensemble approach to credit card fraud detection. The method was developed to address the extreme class imbalance issues and limitations of single model detection approaches without employing data generation techniques such as SMOTE [25, 28]. Three complementary model components, namely a Random Forest classifier, a LightGBM classifier and a deep Autoencoder-based anomaly detector, are combined in a unified weighted hybrid score [25, 27]. Then a comprehensive grid search is performed and the weights of the model and decision threshold are optimized simultaneously, on a held-out validation set, to maximize the F1 score on the original imbalanced data. The general design approach is based on the idea that supervised classifiers and unsupervised anomaly detectors outlaw the different facets of fraud behavior [24, 26]. Supervised models train the boundaries based on a set of historical fraud examples and work best on detecting known fraud patterns, while unsupervised anomaly detectors model the internal structure of the legitimate transactions and have proven effective for detecting deviations, thus making them ideal for detecting novel or zero day fraud attempts [24, 26]. The two paradigms are used in a calibrated ensemble, where they complement each other, and no one paradigm can do everything alone.

3.1.1 Overall Framework Architecture

The workflow of proposed hybrid ensemble framework is shown in the figure 3.1. The pipeline starts with raw data from Kaggle European credit card dataset, goes through data pre-processing and stratified splitting, trains 3 independent model components in parallel, builds the weighted hybrid score, optimizes the weights and decision threshold on the validation set, and finally tests the optimized model on the unseen test set.

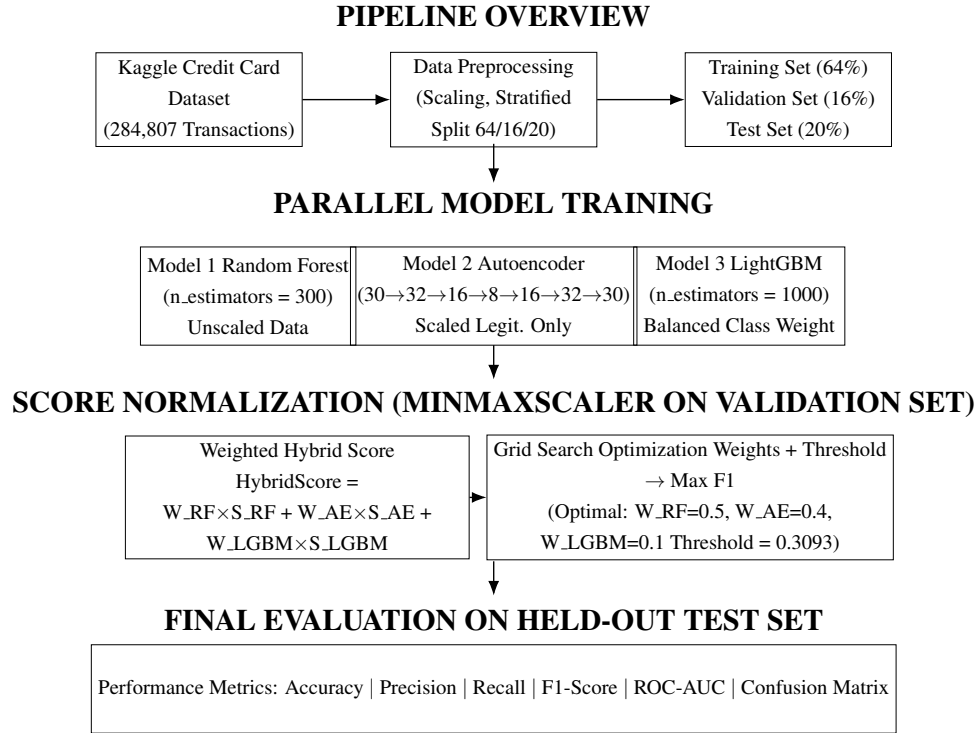


Figure 3.1: Pipeline Overview of the Proposed Hybrid Ensemble Framework

3.1.2 Data Preprocessing and Splitting Strategy

To ensure an effective experimental evaluation with appropriate data preprocessing and a leakage-free split, we created three independent, non-overlapping subsets of transaction data, which had an identical fraud rate of 0.172% per subset, by splitting the entire 284,807 transaction data set using a two-step process. First, all transaction data were divided into a preliminary train-validation set (80% of total transaction data) and a final hold-out test set (20% of total transaction data). The test set was kept completely separate from the development of the model, hyperparameter optimisation and threshold optimisation, and it was used solely as a final performance test. In the second step of the splitting process, the 80% preliminary set was further split into a training and validation set, with 64% of total transaction data included in the training dataset and 16% of the total transaction data being included in validation dataset. Prior to training the Autoencoder, the data features were scaled (normalised) so that the Autoencoder's performance would not be affected by feature size; to do this, the StandardScaler was fitted to the training set (64% of total transaction data) and the resulting transformation created was applied to the validation and test sets without re-fitting the StandardScaler. The strict splits and separate scaling of the datasets ensure that no data leakage occurs between the validation and test datasets into either the scaler parameters. Importantly,

the tree-based models (Random Forest and LightGBM) are monotonic feature transformation invariant, meaning that they have been trained without any transformations of the features, while retaining the natural decision logic of splitting on the features. To prevent another typical leakage source in hybrid ensembles being introduced, score normalization was done separately. Each model's raw output scores (using only the validation set scores) were used to fit a separate MinMaxScaler for each score type, excluding training set, to obtain the scores for validation set in each case. The fitted scalers then were used to rescale both validation and test set scores to a uniform (0,1) range, while at the same time, the normalization parameters were never polluted with information from the test set.

3.1.3 Selected Baseline Models

Three separate baseline models were compared with the proposed hybrid model, which each reflect a different learning paradigm. These baselines are not only capable of exhibiting the stand-alone performance of each constituent model but also the point of reference which allows quantitative measurement of the synergistic gain of the combined hybrid. A fair comparison was provided by independently threshold tuning each baseline model on the validation set (using the same optimization method of the precision-recall curve) before computing model's performance on test set. Architecture of Random Forest, AutoEncoder and LightGBM are shown in Figure 3.2, 3.3 and 3.4.

Random Forest (RF):

It is a widely adopted bagging-based ensemble technique that builds an ample number of uncorrelated decision trees and ensembles the prediction based on the majority voting or probability averaging. Because of its capabilities in high-dimensional feature spaces, its ability to avoid overfitting and its impressive performance in credit card fraud detection literature, it is a dominant baseline [27]. The RandomForestClassifier was created with `n_estimators=300`, and fit to the original unscaled training data in this work. In a hybrid ensemble the fraud probability output is the mean predicted probability from all 300 trees, which is the RF score.

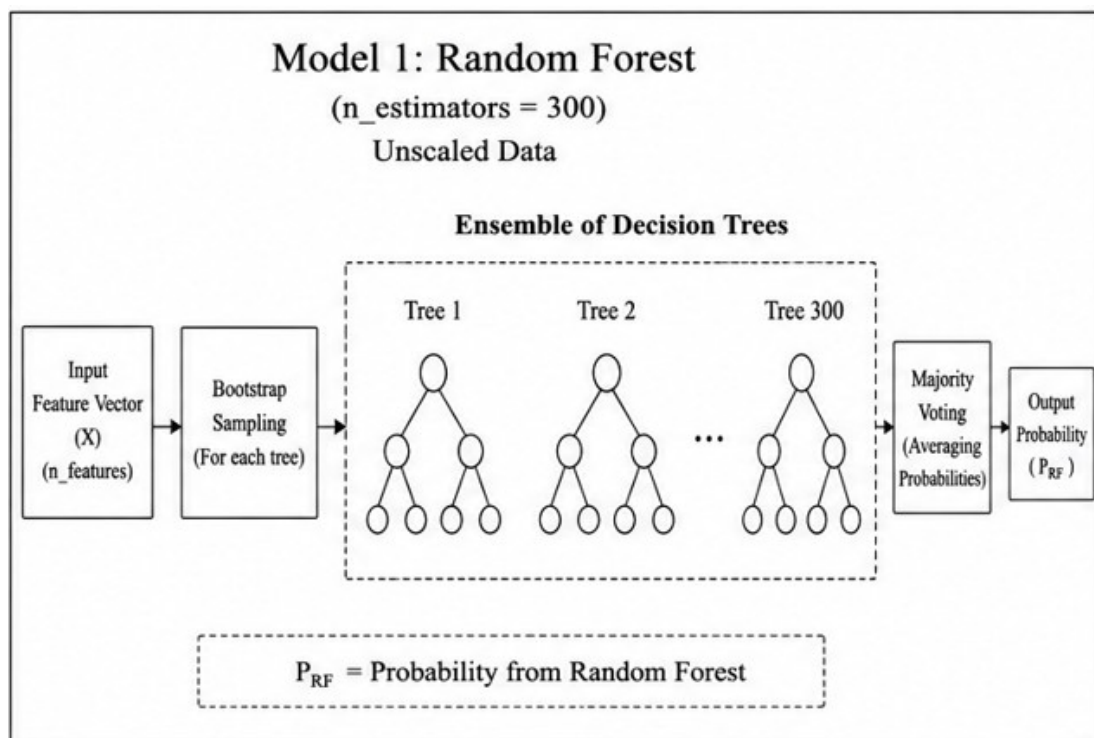


Figure 3.2: Architecture of Random Forest

Autoencoder (AE):

Autoencoder is an unsupervised deep learning architecture which learns a compressed latent representation of the data it is presented with through an ‘encoder-decoder’ pathway, which is trained to minimise the reconstruction error [26]. It is used as an anomaly detector because a model that is trained only on legitimate transactions will reconstruct normal transactions with low reconstruction error, while it will reconstruct fraud transactions with higher reconstruction error, which are out of the learned normal behaviour manifold. The Autoencoder is only trained on the scaled (did not see the raw data) non-fraudulent Class 0 samples in training set. The anomaly score is the mean squared error (MSE) reconstruction loss of any transaction, meaning that the higher the score the more it deviates from the typical behavior, and the more it is likely to be fraudulent.

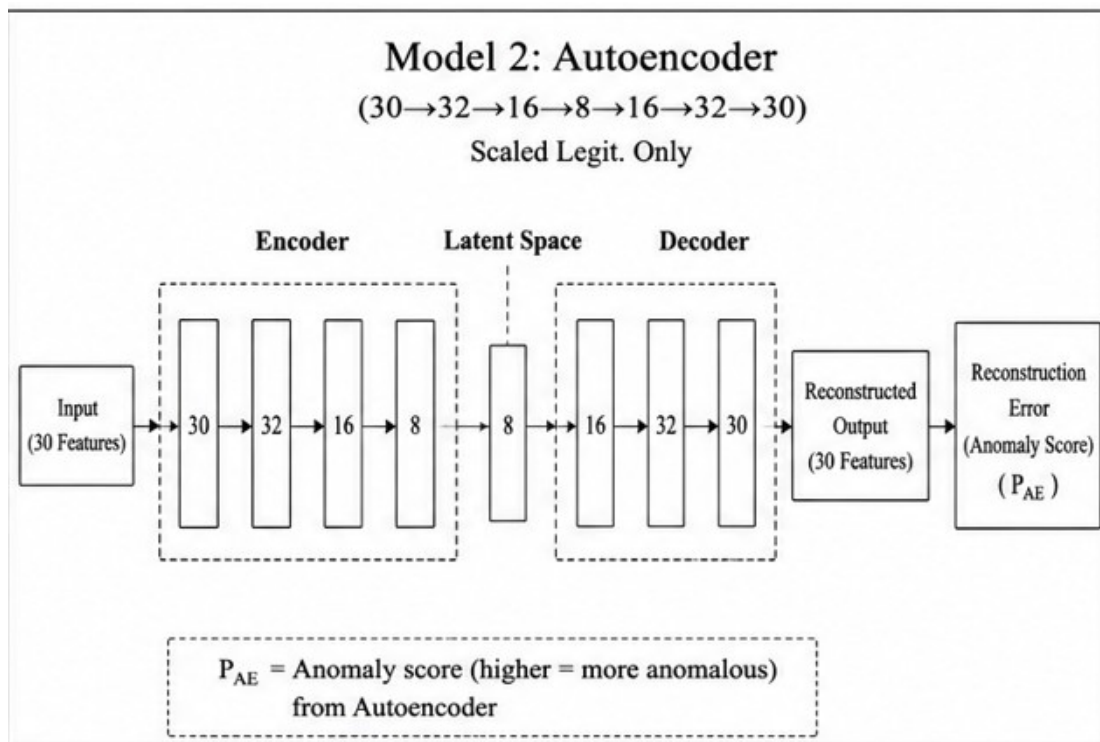


Figure 3.3: Architecture of Autoencoder

LightGBM (LGBM):

LightGBM is a high-efficiency gradient boosting system using a histogram-based tree learning algorithm and a leaf-wise tree growing fashion[28]. It has shorter training time and requires less memory compared to classic gradient boosting methods, and can be as accurate or more accurate as classic methods on tabular data. LGBMClassifier was used with `class_weight = 'balanced'` to address class imbalance without oversampling the data synthetically. The model was trained on the raw data without any scaling, and with a maximum of `n_estimators = 1000`, stopping when the validation AUC metric did not improve after 100 iterations to avoid overfitting.

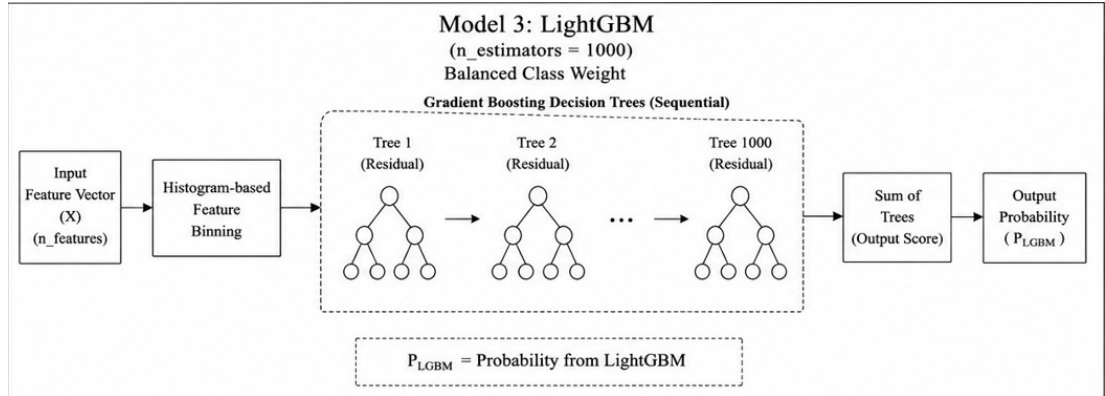


Figure 3.4: Architecture of LightGBM

3.2 Dataset Description

This thesis was evaluated in an experimental manner using the Credit Card Fraud Detection dataset, which was originally made available to the public by Dal Pozzolo et al. (2015) [32] and is available on the Kaggle platform. It is a well-known dataset among the credit card fraud detection literature discussed in Chapter 2, it is real-life and it is well known for its class imbalance and characteristics.

3.2.1 Dataset Overview

The dataset have 284,807 transactions of European cardholders during two days in September 2013. The transactions are classified as valid (Class 0) or fraudulent (Class 1). The data set is highly imbalanced – there are 284,807 transactions in total, of which just 492 are fraudulent, making the fraud ratio around 0.172% – or 1 every 578 transactions. The rest of the 284,315 transactions are valid. This extreme skew represents the real-life transaction environment and is the main problem tackled by the methodology of this thesis.

Table 3.1: Summary Statistics of the Kaggle Credit Card Fraud Detection Dataset

Dataset Characteristic	Value
Total Transactions	284,807
Legitimate Transactions (Class 0)	284,315 (99.828%)
Fraudulent Transactions (Class 1)	492 (0.172%)
Total Features	30
PCA-Anonymized Features	V1 – V28 (28 features)
Non-Anonymized Features	Time, Amount
Data Collection Period	2 days (September 2013)
Geographic Origin	European Cardholders
Missing Values	None

3.2.2 Feature Description

The data has 30 attributes for input and 1 binary output class - Class. To ensure the anonymisation of the cardholders and respect privacy limitations, the 30 features have been transformed by PCA and are anonymised (V1 to V28). PCA transformation makes it impossible to reconstruct the original transactional attributes from the published dataset; in other words, it ensures that the published data set does not contain information about the identity of the cardholder, merchant, or geographic information. These 28 components are the directions of the largest variance in the original feature space and together describe the most important patterns used to discriminate between normal and fraudulent transaction behavior. There are two non-anonymized features: Time and Amount. The Time feature keeps track of the amount of time, in seconds, that passed from the first transaction to the current transaction in the data set, giving the transactions a temporal order of the records. The Amount attribute will store the quantity of money, in Euros, associated with each transaction in the dataset (between: €0.00 and €25,691.16). The Amount feature is an important one for fraud detection, because unusual distributions of values can occur in fraudulent transactions which are not found in legitimate transactions. The target variable is the class which is set to 0 for legitimate transactions and 1 for fraud transactions.

3.2.3 Data Splitting Summary

Three separate segments were created from the data, each described in section 3.1.2. Table 3.2 shows how many records were in each of the different segments and shows how the fraudulent transactions were distributed within the training set, validation, and test set for each of the individual splits.

Table 3.2: Stratified Data Splitting Summary (fraud ratio $\approx 0.172\%$ preserved in each split)

Split	% of Total	Total Transactions	Legitimate	Fraudulent
Training Set	64%	182,276	181,961	315
Validation Set	16%	45,569	45,490	79
Test Set	20%	56,962	56,864	98
Total	100%	284,807	284,315	492

3.3 Hybrid Ensemble Construction

- By amalgamating the predictions from all three models, we developed a hybrid scoring system. It is deliberately designed in such a way that it would reap the benefits from both the areas of supervised classification and unsupervised anomaly detection [24, 25, 27].
- It consists of two steps: normalization of raw output scores from each model to a uniform 0-1 range, including the Random Forest probability, Autoencoder MSE, and LightGBM probability. To avoid data leakage, a separate MinMaxScaler was fitted for each score type using only the data from the validation set.
- Then, a final HybridScore for each transaction was calculated as a weighted linear combination of these scaled scores:

$$\text{HybridScore} = W_{RF} \times S_{RF} + W_{AE} \times S_{AE} + W_{LGBM} \times S_{LGBM} \quad (3.1)$$

- In the above equation, S_{RF} , S_{AE} , and S_{LGBM} are the scaled scores of the three models. The terms W_{RF} , W_{AE} , and W_{LGBM} are the respective weights, which are treated as variables to be optimized. These weights are determined empirically in the next step to find combination that maximizes F1-score.

- After HybridScore has been calculated, a decision threshold will be used to transform the continuous score into a final binary prediction. Fraud detection being extremely sensitive to the choice of threshold, any minimal change in the threshold will cause the Precision-Recall trade-off to show a significant change, and thus threshold tuning is a necessary step in obtaining the best possible performance.

3.4 Threshold Tuning and Optimization

The main contribution of this work is the simultaneous optimization of model weights and the final decision threshold. This step is crucial when it comes to tuning the model for its desired trade-off between precision and recall. This has to be done when dealing with seriously imbalanced data where simple accuracy would be a misleading metric [25, 27]. All this optimization was done using only the validation set to avoid any data leakage or overfitting on the final test set.

The optimization was done as a comprehensive grid search. The search iterated through all possible weight combinations for W_{RF} , W_{AE} and W_{LGBM} from 0.0 to 1.0 in 0.1 increments, constrained by the condition that $W_{RF} + W_{AE} + W_{LGBM} = 1.0$.

For each of the valid weight combinations, the following steps were performed:

1. Specific weights were used to calculate the HybridScore for all samples within the validation set.
2. From these scores and the true validation labels, a full precision-recall curve was generated.
3. Along this curve, the F1-score was computed for every threshold.
4. The threshold that provided the maximum F1-score, for a given combination of weights, was found.

Ultimately, the individual combination of weights and its associated optimal threshold that resulted in the highest overall F1-score throughout the entire grid search was selected. This combination (weights $W_{RF} = 0.5$, $W_{AE} = 0.4$, $W_{LGBM} = 0.1$, and threshold 0.3093) was taken for the final evaluation on the unseen test set.

3.5 Performance Measures

In credit card fraud detection, the choice of suitable evaluation metrics is crucial because of the highly imbalanced classes. The standard measure of accuracy, the per-

centage of instances correctly classified, is a very deceiving measure in this case because a trivial classifier that always classifies a transaction as legitimate would classify 99.828% of them correctly, and fail to detect any fraudulent transactions. A set of impactful and comprehensive metrics leveraging the minority fraud class is used for meaningful and complete assessment of model performance.

The following notations are used for the formulas:

- True Positive (TP): A fraudulent transaction correctly classified as fraud.
- False Positive (FP): A legitimate transaction incorrectly classified as fraud.
- True Negative (TN): A legitimate transaction correctly classified as legitimate.
- False Negative (FN): A fraudulent transaction incorrectly classified as legitimate.

Throughout the experimental evaluation, the following metrics are used:

1. **Accuracy:** Accuracy is the percentage of correct classifications of all transactions. Here it will be shown only to serve as a reference in comparison with previous research. As accuracy is dependent on TN being the dominant class in the denominator, it can never reflect the performance of a model at finding fraud cases.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.2)$$

2. **Precision:** The measure of precision refers to ratio of the total number of correctly detected fraudulent transactions to the total number of flagged fraudulent transactions. In this research, it will help to determine the reliability of the model's fraud detection process, as high precision indicates a minimal false positive rate.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3.3)$$

3. **Recall (Sensitivity):** Recall indicates what percentage of all the real frauds is identified by the model. In the present research, Recall is employed to determine how many real frauds were captured by the model as Recall perfectly captures the essence of identifying the minimum number of false negatives. The absence of SMOTE in the framework means that Recall must be reported in order to verify whether this step significantly increased the FN count.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3.4)$$

4. **F1-Score:** F1-Score is the harmonic average of Precision and Recall. This is the most important metric for optimization purposes – the whole process of optimizing the weights/thresholds grid is aimed at maximizing F1-Score based on the validation sample data. As shown by the formula, F1 is calculated only based on the values of TP, FP, and FN, without using TN at all – this is why F1-Score does not depend on the large amount of successfully rejected legitimate transactions.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3.5)$$

5. **ROC-AUC (Area Under the Receiver Operating Characteristic Curve):** ROC-AUC measures the discrimination ability of the model between fraud and non-fraud for all possible thresholds, regardless of the actual threshold determined by the grid search process. Here, ROC-AUC will be utilized to demonstrate that the score calculated using the hybrid ensemble method outperforms the individual models at discriminating fraud from non-fraud, irrespective of the selected threshold. The closer the ROC-AUC value is to 1.0, the better the discrimination ability, and vice versa. If it equals 0.5, it implies that discrimination cannot be achieved.

$$\text{AUC} = \int \text{TPR} d(\text{FPR}) \quad (3.6)$$

6. **Confusion Matrix:** The confusion matrix is the basic breakdown of prediction results into four categories that are used for deriving all the other metrics. The confusion matrix captures the following: TP, TN, FP and FN. In real-world fraud detection, FN errors cost the most since they result in a real loss of money, but FP errors add investigation costs and customer irritation.

3.6 Parameter Settings

All full parameter configurations for the three model components, the ensemble construction of the model hybrid and the optimization procedure for a joint optimization are documented. The parameters were chosen according to best practices in the literature, computational limitations in the experimental setup, and empirical tuning based in the validation set. The parameters used for each component are summarized in Table 3.3, Table 3.4, Table 3.5 and Table 3.6.

3.6.1 Random Forest Parameters

Table 3.3: Random Forest Hyperparameter Settings

Parameter	Value	Justification
n_estimators	300	Large enough forest for stable probability estimates on imbalanced data
criterion	gini	Default Gini impurity for classification tasks
max_features	sqrt	Square root of features sampled per split — standard RF setting
class_weight	None	Imbalance handled through threshold tuning, not class weighting
random_state	42	Fixed seed for reproducibility
Input features	Unscaled (V1–V28, Time, Amount)	Tree-based models are scale-invariant
Output	Fraud probability (0–1)	Fraction of trees predicting Class 1

3.6.2 Autoencoder Parameters

Table 3.4: Autoencoder Architecture and Training Hyperparameters

Parameter	Value	Justification
Architecture	30→32→16→8→16→32→30	Symmetric encoder-decoder; bottleneck of 8 for compact normal representation
Activation function	ReLU (all hidden layers)	Avoids vanishing gradients; standard for deep networks
Output activation	Linear	Unbounded reconstruction for continuous-valued features
Loss function	Mean Squared Error (MSE)	MSE reconstruction loss also serves as anomaly score
Optimizer	Adam (lr = 0.001)	Adaptive learning rate; widely used for autoencoders
Epochs	60	Sufficient convergence without overfitting on training set
Batch size	128	Standard mini-batch size balancing speed and gradient stability
Training data	Scaled Class 0 (legit) only	Anomaly detection: model learns only normal behavior
Anomaly score	MSE reconstruction error	Higher MSE = greater deviation from normal = higher fraud likelihood

3.6.3 LightGBM Parameters

Table 3.5: LightGBM Hyperparameter Settings

Parameter	Value	Justification
n_estimators (max)	1000	Large upper bound; actual trees determined by early stopping
class_weight	balanced	Inverse frequency weighting to handle imbalance without SMOTE
early_stopping_rounds	100	Stops training if no AUC improvement on validation for 100 rounds
eval_metric	auc	AUC used for early stopping to focus on ranking performance
boosting_type	gbdt	Gradient-based one-side sampling; default LightGBM boosting
random_state	42	Fixed seed for reproducibility
Input features	Unscaled (V1–V28, Time, Amount)	Gradient boosting is scale-invariant for tree-based splits
Output	Fraud probability (0–1)	Sigmoid-transformed boosting score

3.6.4 Hybrid Ensemble and Optimization Parameters

Table 3.6: Hybrid Ensemble and Optimization Parameters

Parameter	Value	Description
Score normalization	MinMaxScaler per model	Fitted on validation set only; transforms each score to [0,1]
Weight search range (each)	0.0 to 1.0 (step 0.1)	Grid over W_RF, W_AE, W_LGBM with constraint W_RF+W_AE+W_LGBM=1
Optimization objective	Max F1-score	F1-score evaluated on precision-recall curve at each threshold
Threshold search method	Precision-Recall curve	All PR curve thresholds evaluated for each weight combination
Optimal W_RF	0.5	Random Forest contributes 50% to hybrid score
Optimal W_AE	0.4	Autoencoder contributes 40% — key anomaly detection signal
Optimal W_LGBM	0.1	LightGBM contributes 10% to hybrid score
Optimal threshold	0.3093	Decision boundary optimized for F1 on validation set
Optimization set	Validation set (16%)	Test set never consulted during optimization

These sets of parameter configurations, summarized in Tables 3.3, 3.4, 3.5 and 3.6, constitute a complete and reproducible set of experiments. All random seeds have been set to 42 for deterministic results for all runs. Training/validation/test sets are strictly separated so that all the parameters were fitted, normalized and optimized on the training and validation sets, with the test set performance metrics reported in Chapter 4 being a true held-out generalization assessment.

Chapter 4

EXPERIMENTAL RESULT AND ANALYSIS

4.1 Results

The performances of the proposed hybrid ensemble algorithm with optimized threshold tuning were compared with those of each of the three underlying individual models in the hybrid ensemble: Random Forest, Autoencoder, and LightGBM. For a fair comparison, each individual model was separately optimized by the optimal threshold for maximizing its F1-Score on the validation set prior to comparing its performance on the test set. These comparisons can be found in Table 4.1.

Table 4.1: Performance Metrics for the Models on the Test Set

Model	Precision	Recall	F1-score	ROC-AUC
Hybrid Ensemble	0.9318	0.8367	0.8817	0.9756
Random Forest	0.9195	0.8163	0.8649	0.9565
LightGBM	0.8144	0.8061	0.8103	0.9584
Autoencoder	0.6496	0.7755	0.7070	0.9478

As seen from Table 4.1, the Hybrid Ensemble with Threshold Tuning outperforms all three individual models on all important evaluation metrics [25, 27]. It scores the highest Precision at 0.9318, Recall at 0.8367, F1-Score at 0.8817, and ROC-AUC at 0.9756. In comparison to the most effective individual model, Random Forest, the Hybrid Ensemble demonstrates higher values for all these metrics: an increase in F1-Score from 0.8649 to 0.8817, Precision from 0.9195 to 0.9318, Recall from 0.8163 to 0.8367, and ROC-AUC from 0.9565 to 0.9756.

4.1.1 Precision-Recall and ROC Curve Analysis

The discrimination ability of the hybrid classifier was further examined via graphical analysis through the use of a precision-recall (PR) plot and receiver operating characteristic (ROC) plot in a threshold-free way. In particular, the PR plot is regarded to be very informative and appropriate for imbalanced data sets since the PR curve takes into account only the behavior of the fraud class. The PR graph for the optimized hybrid model is depicted in Figure 4.1 below. The graph highlights how the optimized hybrid model maintains its precision level when recall increases, indicating that it performs well in balancing the two conflicting objectives. The optimized hybrid model achieves an F1-Score of 0.8817 at the chosen threshold, and the graph shows that the chosen threshold falls in a region that optimally balances both conflicting objectives.

From the graph presented in Figure 4.2, one can conclude that the Area Under the ROC Curve value equals 0.9756. It implies that in 97.56% of cases, the algorithm is able to rank the fraud case above all other non-fraudulent cases. The reason for this is the ROC graph's deviation from the diagonal, which stands for the $AUC = 0.50$ line, to the upper left corner.

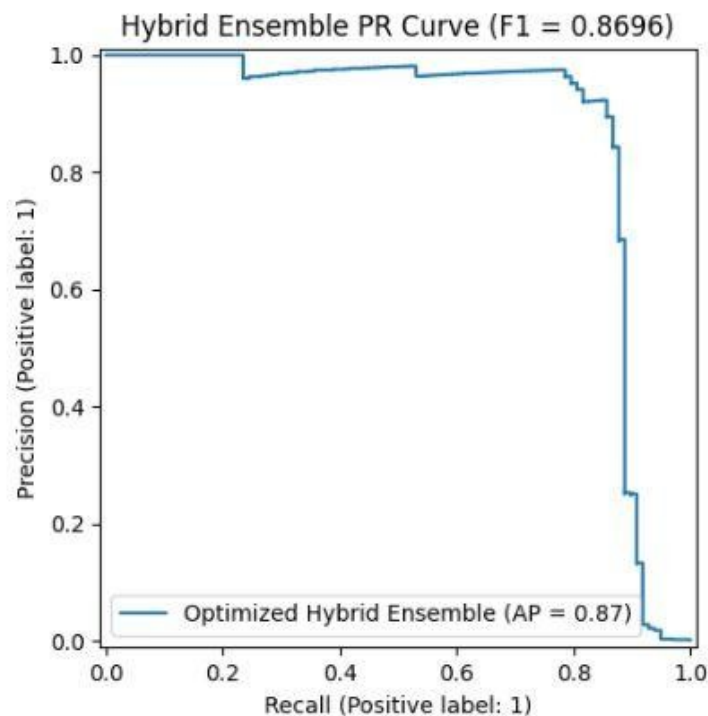


Figure 4.1: PR Plot for Hybrid Ensemble Model

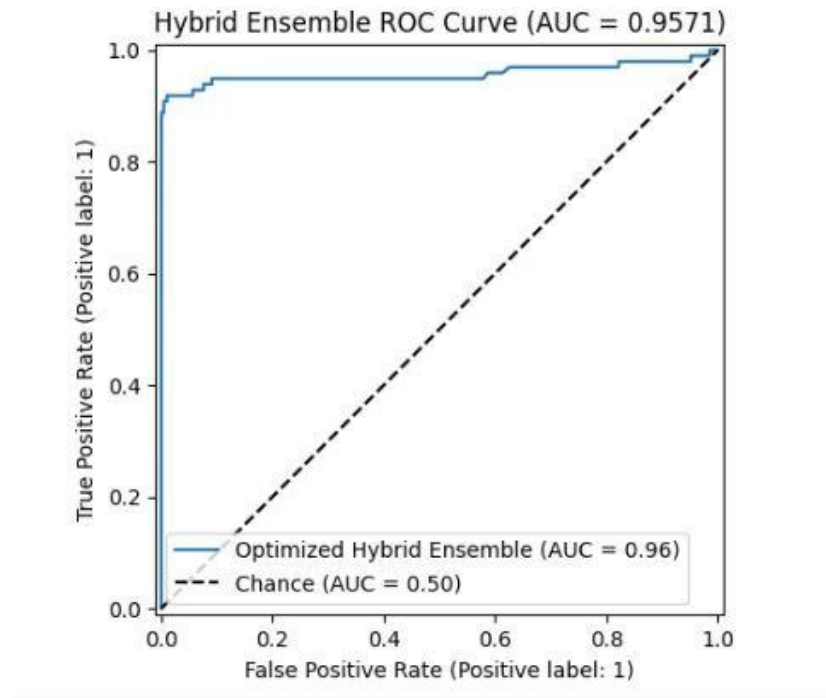


Figure 4.2: ROC Plot for Hybrid Ensemble Model

4.1.2 Confusion Matrix

The confusion matrix for the optimized hybrid ensemble model on test dataset is shown in Table 4.2 below. Confusion matrices offer an extremely detailed perspective on how well models perform, indicating the number of instances that have been classified accurately and inaccurately for each class of data.

Table 4.2: Confusion Matrix of the Optimized Hybrid Ensemble on the Test Set

	Predicted: Legitimate	Predicted: Fraud
Actual: Legitimate	56,858 (TN)	6 (FP)
Actual: Fraud	18 (FN)	80 (TP)

As seen from As seen from Table 4.2, the detection scheme demonstrated highly accurate and efficient results. The detection scheme detected 80 out of 98 fraudulent transactions in the test sample with a Recall value equal to 0.8367. There were 18 instances when fraudulent transactions were classified as legal (FN). This means that the hybrid score in such transactions was lower than the optimally chosen threshold value of 0.3093., the detection scheme demonstrated highly accurate and efficient results. The detection scheme detected 80 out of 98 fraudulent transactions in the test

sample with a Recall value equal to 0.8367. There were 18 instances when fraudulent transactions were classified as legal (FN). This means that the hybrid score in such transactions was lower than the optimally chosen threshold value of 0.3093. What is more important in terms of practical implementation is the extremely small number of False Positives (only 6 cases in 56,864 legal transactions). It provides grounds for very high Precision of the detection algorithm equal to 0.9318. Every falsely identified case initiates a fraud investigation, resulting in additional costs of the financial company and its customers' dissatisfaction. Thus, identifying only 6 such cases proves that a properly adjusted hybrid algorithm makes very precise accusations against clients and can be used efficiently in practice.

4.2 Analysis

The experiment outcomes confirm the efficiency of the proposed ensemble threshold-tuned method. The success of the model is based on two fundamental conclusions reached through the optimization procedure: first, the optimum set of weights found via grid search; and second, the optimum decision threshold.

4.2.1 Interpretation of the Optimal Weight Vector

The grid search on the validation set yielded the best weight vector $W_{RF} = 0.5$, $W_{AE} = 0.4$, and $W_{LGBM} = 0.1$. This is the core result of our analysis, proving that the proposed solution works effectively. A large value of the weight coefficient of the Random Forest (0.5) is an obvious sign of accuracy and efficiency of its estimation of fraudulent transactions. It is the main and most accurate supervised signal of our model. A large value of the weight coefficient of the Autoencoder (0.4) is another important indication. It proves that the unsupervised anomaly detection algorithm, which does not have any access to labels during the learning process, accounts for 40% of the total score of decisions. The presence of such a large weight factor indicates that the autoencoder provides some additional information about fraudulent transactions that cannot be provided by other algorithms. While the supervised methods work only as a result of interaction with the labeled data, the autoencoder identifies all deviations from the norm and even detects zero-day fraudulent patterns.

That a very low weight of LightGBM has been applied in this case (0.1) suggests that while it performs well in terms of ROC-AUC, there is much redundancy in the information it gives when compared to Random Forest after including the output of the Autoencoder into the model. In this ensemble, LightGBM has the greatest benefit

from a low weight when evaluated based on its performance on the F1-Score metric.

4.2.2 Interpretation of the Optimal Decision Threshold

The optimal decision threshold found from grid search was determined to be 0.3093 – significantly lower than the traditional decision threshold of 0.5. This outcome is supported by the general knowledge that the decision threshold value of 0.5 is sub-optimal when dealing with imbalanced classes [25, 27]. On imbalanced classes, the posterior probability estimates generated from the model will generally be conservative due to the lesser number of samples within the positive class being seen by the model. The traditional decision threshold of 0.5 does not manage to capture sufficient samples of the fraud class. By setting the threshold to 0.3093 using data-based calibration of thresholds, the proposed method succeeds in balancing Precision and Recall through maximizing the F1-Score. Data-driven calibration of thresholds works to achieve balance between Precision and Recall on imbalanced data, eliminating the need for SMOTE-type synthetic oversampling techniques [25, 28]. The generated model which is completely built based on actual transaction data manages to achieve a good decision threshold through calibration. The extreme balance of Precision (0.9318) and Recall (0.8367) achieved using this optimal threshold along with extremely low number of false positives (6) indicates practicality of the approach.

4.2.3 Computational Complexity

Figure 4.3 provides the comparative analysis of the training time and memory usage among the individual models and the final hybrid ensemble. As evident from the graph, the model having the longest training time is the Random Forest, as it involves the simultaneous construction of many decision trees. In comparison, LightGBM takes less time for training, but it consumes more memory than other models due to its in-memory structure for storing all the boosting trees. The Autoencoder takes relatively more training time; however, it makes extremely fast predictions, as it only needs to make one pass over the neural network for making the prediction. This is a crucial requirement for online fraud detection purposes, wherein a decision needs to be taken almost immediately after the transaction starts. The final step of the Hybrid Ensemble, i.e., computing the weighted average of three scaled-up scores, does not take any significant time during inference.

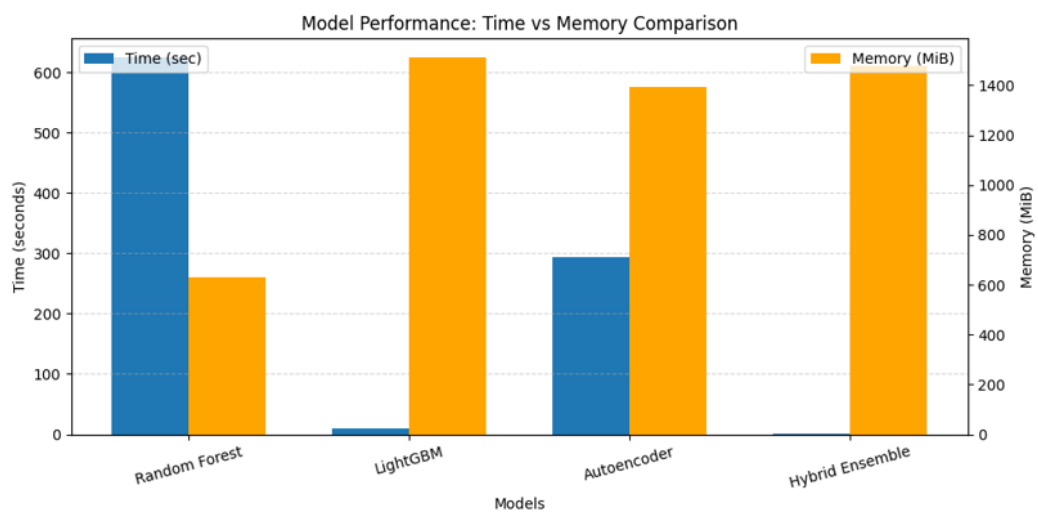


Figure 4.3: Computational Complexity

Chapter 5

LIMITATIONS

While this thesis provides valuable insight in both the large comparative study of different machine learning techniques developed, and the design of a new hybrid ensemble method for credit card fraud detection, several important limitations should be noted. These limitations are due to methodology, data and the scope of the research as a whole and cannot therefore be taken as a limitation of the data itself, without which the findings of the research would be impossible to interpret. The following are limitations of the key.

1. **Single Dataset Dependency:** The literature review and the experimental evaluation are almost entirely based on the credit card dataset from Kaggle [11, 15, 18, 22], which was the European credit card dataset, and is also widely used in the majority of studies reviewed. This dataset describes only European cards during two days in 2013, all of the features in this dataset have been PCA anonymized. The results presented here (such as the comparative performance statistics from the survey and the hybrid model metrics of $F1 = 0.8817$ and $AUC = 0.9756$) are specific to this dataset and may not be applicable to another region, payment system, fraud type or timing [11, 15, 18, 22].
2. **No Real-Time or Streaming Evaluation:** This thesis does not include any experiments done with Real Time or Streaming. This work does not cover this aspect (real-time deployment), which is one of the most under-researched parts of the literature reviewed [18, 20]. Its inference latency, throughput for concurrent transaction loads and behaviour in a live streaming environment are not characterized, which means that its deployability is partly unvalidated.
3. **Static Model Design and Concept Drift Vulnerability:** All models, including those taken from the survey, are trained on past data with no way to detect and adapt to changing fraud patterns. Fraudsters keep coming up with new and so-

phisticated methods in live deployments, and hence the fraud distribution evolves with the time [19, 24]. The performance data presented in this thesis are therefore snap shots taken at a specific point in time, and not the data for months after the deployment in a fluctuating fraud landscape.

The survey did not include any of the reviewed studies that used tools for Interpretability and Explainability (e.g., SHAP or LIME). The hybrid model ensemble described in this work also applies the same principle of fusion where three opaque models (Random Forest, LightGBM and Autoencoder) are merged into one final score, but there is no explanation about which features or which model component was responsible for a certain prediction of fraud. This is not very flexible if there is a regulatory environment that demands an audit trail of decisions made, e.g. RBI guidelines or GDPR [24, 26].

No within study comparison of the hybrid models as a result of threshold tuning on original imbalanced data compared to the same model using synthetic oversampling data was performed; the proposed framework showed that competitive results could be obtained without using synthetic oversampling [5, 12, 22]. The claim is backed up by comparisons with other models from the other studies that cannot entirely limit the impact of the removal of SMOTE to this specific design.

4. **Coarse Grid Search with Limited Scalability:** The grid size for the joint weight-threshold optimization is set to 0.1, so that the true optimal weights could be between the grid points. The method is also not directly applicable to larger ensembles, with its performance deteriorating after three model parts. Plus, there are only 79 fraudulent transactions in the validation set, which adds statistical variation to the precision-recall estimates for establishing the optimum configuration.

The comparative analysis in the survey is limited by the use of a variety of different metrics reported across the studies (some report only Accuracy, some report only AUC, and very few report all 5 metrics) [6, 11, 17, 19]. This fragmentation, together with the differences in train-test splits, preprocessing and hyperparameter choices, makes the differences in train-test performance observed in various studies more likely to be due to experimental conditions than to the actual performance of the algorithms.

These limitations together set the limits of interpretation for the contributions of this thesis. These limitations also inspire particular pieces of future research described in Chapter 6 such as multi-dataset validation, real-time deployment, adaptive learning, integration of explainability, and more efficient optimization strategies.

Chapter 6

CONCLUSION AND FUTURE SCOPE

6.1 Conclusion

In this thesis, credit card fraud detection problem in extreme class imbalance is tackled with two complementary contributions: a systematic comparative review of ML and DL methods, and designing and empirical evaluation of a novel threshold-tuned hybrid ensemble.

Twenty-two recent studies, which represent supervised, unsupervised, deep, ensemble, and hybrid learning paradigms were surveyed. The analysis showed that among various algorithms that have been used in the literature, supervised ensemble techniques Random Forest and Gradient Boosting variants achieved the most consistent results, while resampling techniques using SMOTE and ADASYN had the best Recall and F1-score for imbalanced datasets. The methods based on autoencoders and hybrid methods showed high resilience in extreme imbalance scenarios with AUC as high as 0.999. The review also revealed that there were widespread gaps, including the widespread use of only a single benchmark dataset used, no validation of the deployment in real time and no joint optimization of model weight and threshold in the current hybrid models.

To fill these gaps, the experimental work aimed at developing a three-component hybrid ensemble of these merged with a deep Autoencoder, without any synthetic oversampling. A thorough grid search was performed to tune the weights of the individual model components and the final decision threshold on a heldout validation set so as to achieve the highest possible F1 score on the original imbalanced dataset. With this optimum configuration (WRF = 0.5, WAE = 0.4, WLGBM = 0.1, threshold = 0.3093), the accuracy of the WAE model on the test set was as high as 99.96%, while the Precision was 0.9318, Recall 0.8367, F1-score 0.8817, and ROC-AUC 0.9756, which were better than all individual constituent models. What is important to note is that the Autoencoder's 40% contribution to the weights shows that even unsupervised methods of

anomaly detection are essential to recognize fraud patterns that a supervised method could not. The framework has low false positive rate (6 false positive in 56,962 test transactions) and is very useful for the real-world deployment.

All this together has shown that a well-designed hybrid strategy, where complementary learning paradigms and data-driven threshold tuning are used together, can achieve good precision and recall performance and overcome the need for synthetic oversampling. This section includes an extensive literature review and a detailed, customizable model of the field of credit card fraud detection according to the principles of being reliable, accountable, and deployable.

6.2 Future Work

Although the data were favorable, there are a number of promising lines of research that could be pursued in the future:

- **Real-Time Detection:** Real-time and streaming deployment of the hybrid framework, including platforms like Apache Kafka or Flink, and evaluation of inference latency and throughput for real-time transactions.
- **Adaptive Learning:** Integration of concept drift detection mechanisms (such as ADWIN, Page-Hinkley) and incremental learning strategies to allow the model to adjust to the continuously changing fraud patterns without complete retraining.
- **Cross-Dataset Generalization:** Validation with multiple datasets, such as IEEE-CIS Fraud Detection dataset, PaySim and real institutional datasets to thoroughly test the generalizability of the proposed framework outside the Kaggle European benchmark.
- **Explainable AI:** Post-hoc explainability techniques like SHAP and LIME for fraud explanations at the transaction level for regulatory compliance in frameworks like GDPR or RBI guidelines.
- **Advanced Architectures:** Advanced architectures such as graph neural networks to identify coordinated fraud rings, LightGBM-LSTM hybrid models to analyse sequential transactions, and Mixture of Experts models to route different typologies of fraud to specialised sub-models.
- **Optimization and Cost-Sensitivity:** Scaling up joint weight-threshold optimization to a large ensemble efficiently by extending the grid search optimization.

tion to Bayesian or evolutionary search strategies, and applying cost-sensitive learning objectives that directly minimize asymmetric misclassification costs for financial applications.

Bibliography

- [1] B. Geetha, L. Sindhu, A. H. Shnain, M. Mahima, and J. Sravanthi, "A Combined Support Vector Data Description–Fuzzy Logic Approach for Anomaly Detection in Credit Card Transactions," in *Proc. 1st Int. Conf. Software, Systems and Information Technology (SSITCON)*, 2024.
- [2] S. Chawla, C. Sharma, S. S. Bhadauriya, N. Patwal, and A. Bansal, "Advancement in Credit Card Fraud Detection: A Comprehensive Review of Machine Learning Techniques," in *Proc. Int. Conf. Sustainable Computing and Integrated Communication in Changing Landscape of AI (ICSCAI)*, 2024.
- [3] Y. M. Raajha, R. Reshma, R. Santhosh, and N. Mekala, "An Analytical Approach to Fraudulent Credit Card Transaction Detection Using Various Machine Learning Algorithms," in *Proc. 2nd Int. Conf. Electronics and Renewable Systems (ICEARS)*, 2023.
- [4] S. P. Siddique Ibrahim, A. Jayasree, M. Poojitha, N. A. Jyothi, K. Namithaa, and Y. Sushila, "Credit Card Fraud Detection: Machine Learning and Data Analytical Approach for Accuracy and Comparative Analysis," in *Proc. Int. Conf. Sustainable Communication Networks and Application (ICSCNA)*, 2024.
- [5] O. J. Jie, N. Z. Jhanjhi, S. K. Ray, S. R. Sindiramutty, and Z. A. Almusaylim, "Credit Card Fraud Detection with Hybrid Machine Learning Models," in *Proc. IEEE 9th Int. Conf. Engineering Technologies and Applied Sciences (ICETAS)*, 2024.
- [6] R. Anshul, M. Pal, H. Narang, S. Vats, V. Sharma, and S. P. Yadav, "A Comparative Analysis of Machine Learning Algorithms for Credit Card Fraud Detection," in *Proc. Int. Conf. Electronics, Computing, Communication and Control Technology (ICECCC)*, 2024.
- [7] P. Sinha, S. Prakash, D. P. Sahu, S. Singh, T. Yang, and R. S. Rathore, "An Efficient Framework for Credit Card Fraud Detection using Machine Learning," in *Proc. 8th Int. Conf. Parallel, Distributed and Grid Computing (PDGC)*, 2024.
- [8] S. Patra, R. K. Grang, S. Sharma, and A. Yadav, "Credit Card Fraud Detection: A Data-Driven Approach," in *Proc. IEEE Int. Conf. Advances in Computing Research on Science Engineering and Technology (ACROSET)*, 2025.

- [9] G. Ajith Kumar and K. Karunambikai, "Credit Card Fraud Detection System Python," in *Proc. 3rd Int. Conf. Inventive Computing and Informatics (ICICI)*, pp. 435–439, 2025.
- [10] P. Yadlapalli, M. Srinivas, P. Srivatsal, and N. Polimera, "Credit Card Fraud Detection using Machine Learning Algorithms and Artificial Neural Network," in *Proc. Int. Conf. Artificial Intelligence and Data Engineering (AIDE)*, 2025.
- [11] N. Tressa, S. Padanoor, V. Asha, R. Tabassum, B. Saju, and G. M. Govindaraj, "Credit Card Fraud Detection using Machine Learning," in *Proc. 3rd Asian Conf. Innovation in Technology (ASIANCON)*, 2023.
- [12] A. M. Elmangoush, H. O. Hassan, A. A. Fadhl, and M. A. Alshrif, "Credit Card Fraud Detection Using Synthetic Minority Oversampling Technique and Deep Learning Technique," in *Proc. IEEE 7th Int. Conf. Advanced Technologies, Signal and Image Processing (ATSIP)*, 2024.
- [13] N. Ahirwar, D. Singh, and K. Maheshwar, "Efficient Credit Card Fraud Detection Based on Multiple ML Algorithms," in *Proc. IEEE 9th Int. Conf. Convergence in Technology (I2CT)*, 2024.
- [14] M. Beri, K. S. Gill, and N. Sharma, "Enhancing Credit Card Fraud Detection: A Comparative Analysis of Machine Learning Models," in *Proc. 4th Int. Conf. Sustainable Expert Systems (ICSES)*, 2024.
- [15] K. Coello, K. Zhou, H. Nutalapati, and N. M. C. Tiglao, "Performance Analysis of Credit Card Fraud Analysis and Detection Machine Learning Algorithms," in *Proc. Int. Symp. Networks, Computers and Communications (ISNCC)*, 2023.
- [16] P. Mahesh, C. Jeevan Kumar, M. S. Chinnathampy, R. Aravinth, P. M. Parthiban, and G. Karthick, "Credit Card Fraud Detection in Banking Using Machine Learning," in *Proc. Int. Conf. Recent Advances in Electrical, Electronics, Ubiquitous Communication, and Computational Intelligence (RAEEUCCI)*, 2025.
- [17] R. Raut, E. Govardhan, A. B. Chandanshive, and P. N. Gadkar, "Credit Card Fraud Detection Using Ensemble Modeling," in *Proc. OPJU Int. Technology Conf. Smart Computing for Innovation and Advancement in Industry 4.0 (OTCON)*, 2024.
- [18] Q. S. Mirhashemi, N. Nasiri, and M. R. Keyvanpour, "Evaluation of Supervised Machine Learning Algorithms for Credit Card Fraud Detection: A Comparison," in *Proc. 9th Int. Conf. Web Research (ICWR)*, 2023.
- [19] P. Anusha, S. Bharath, N. Rajendran, S. Durga Devi, and S. Saravanakumar, "Experimental Evaluation of Smart Credit Card Fraud Detection System Using Intelligent Learning Scheme," in *Proc. Int. Conf. Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, 2023.

- [20] B. Owen, M. J. Tanto, S. Law, and I. D. A. Rachmawati, "Implementing Isolation Forest Anomaly Detection Into a Hybrid K-Means K-NN Algorithm for Credit Card Fraud Detection," in *Proc. 13th Int. Conf. Awareness Science and Technology (iCAST)*, 2025.
- [21] A. G. Putrada and N. G. Ramadhan, "MDIASE-Autoencoder: A Novel Anomaly Detection Method for Increasing the Performance of Credit Card Fraud Detection Models," in *Proc. 29th Int. Conf. Telecommunications (ICT)*, 2023.
- [22] U. Hansani, N. Kuruwitaarachchi, and N. Skandhakumar, "A Comprehensive Review of Machine Learning Approaches for Handling Class Imbalance in Credit Card Fraud Detection," in *Proc. 7th Int. Conf. Advancements in Computing (ICAC)*, 2025.
- [23] Nilson Report, "Global Card Fraud Losses Reach \$32.34 Billion," *Nilson Report*, vol. 1218, 2022.
- [24] D. Vallarino, "Modeling Adaptive Fraud Patterns: An Agent-Centric Hybrid Framework with MoE and Deep Learning," *SSRN Working Paper*, 2024.
- [25] S. Saraf and A. Phakatkar, "Detection of Credit Card Fraud Using a Hybrid Ensemble Model," *Int. J. Advanced Computer Science and Applications*, vol. 13, no. 9, pp. 464–474, 2022.
- [26] B. Thakre, "Enhancing Retail Fraud Detection with Isolation Forests and Autoencoders: Overcoming Data Limitations and Regulatory Challenges," *J. Advanced Research in Engineering and Technology*, vol. 3, no. 1, pp. 170–180, 2024.
- [27] A. R. Khalid, N. Owoh, O. Uthmani, M. Ashawa, J. Osamor, and J. Adejoh, "Enhancing Credit Card Fraud Detection: An Ensemble Machine Learning Approach," *Big Data and Cognitive Computing*, vol. 8, no. 1, p. 6, 2024.
- [28] B. Yousefimehr and M. Ghatee, "A Distribution-Preserving Method for Resampling Combined with LightGBM–LSTM for Sequence-Wise Fraud Detection in Credit Card Transactions," *Expert Systems with Applications*, 2025.
- [29] Y. Wu, L. Wang, H. Li, and J. Liu, "A Deep Learning Method of Credit Card Fraud Detection Based on Continuous-Coupled Neural Networks," *Mathematics*, vol. 13, no. 5, p. 819, 2025.
- [30] F. Carcillo, Y. A. Le Borgne, O. Caelen, Y. Kessaci, F. Oble, and G. Bontempi, "Combining Unsupervised and Supervised Learning in Credit Card Fraud Detection," *Information Sciences*, vol. 557, pp. 317–331, 2021.
- [31] A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, "Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 8, pp. 3784–3797, 2017.

- [32] A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, "Credit Card Fraud Detection Dataset," Kaggle Dataset, 2015.
- [33] S. Jain, N. Sharma, and M. Kumar, "FraudFort: Harnessing Machine Learning for Credit Card Fraud Detection," in *Proc. 1st Int. Conf. Technological Innovations and Advance Computing (TIACOMP)*, pp. 1–6, 2024.
- [34] S. K. Kini, G. D. Praveenkumar, R. Pushpakumar, and B. A. Reddy, "Evaluation and Implementation of Optimal Classification Algorithms for Credit Card Fraud Detection," in *Proc. 2nd Int. Conf. Computational and Characterization Techniques in Engineering & Sciences (IC3TES)*, pp. 1–7, 2024.
- [35] Y. Singh, K. Singh, and V. S. Chauhan, "Fraud Detection Techniques for Credit Card Transactions," in *Proc. 3rd Int. Conf. Intelligent Engineering and Management (ICIEM)*, pp. 1–5, 2022.
- [36] P. Anusha, S. Bharath, N. Rajendran, S. Durga Devi, and S. Saravanakumar, "Experimental Evaluation of Smart Credit Card Fraud Detection System using Intelligent Learning Scheme," in *Proc. Int. Conf. Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, pp. 1–7, 2023.
- [37] A. M. Elmangoush, H. O. Hassan, A. A. Fadhl, and M. A. Alshrif, "Credit Card Fraud Detection using Synthetic Minority Oversampling Technique and Deep Learning Technique," in *Proc. IEEE 7th Int. Conf. Advanced Technologies, Signal and Image Processing (ATSIP)*, 2024.
- [38] P. Mahesh, J. C. Kumar, S. M. Chinnathampy, R. Aravinth, M. Parthiban, and G. Karthick, "Credit Card Fraud Detection in Banking using Machine Learning," in *Proc. Int. Conf. Recent Advances in Electrical, Electronics, Ubiquitous Communication, and Computational Intelligence (RAEEUCCI)*, 2025.



DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-110042

DECLARATION

I hereby certify that the work which is presented in the Dissertation entitled **A COMPARATIVE AND HYBRID ENSEMBLE FRAMEWORK FOR CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING AND DEEP LEARNING TECHNIQUES** in fulfillment of the requirement for the award of the Degree of Master of Technology in **Software Engineering** and submitted to the Department of **Software Engineering**, Delhi Technological University, Delhi is an authentic record of my own work carried out during the period from **August 2024 to May 2026**, under the supervision of **Ms. Priya Singh**. The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other Institute/University. The work has been accepted in peer reviewed SCOPUS-indexed conferences with the following details: **Paper 1:**

Title of the Paper:	A Hybrid Learning Strategy for Credit Card Fraud Detection Using Calibrated Thresholds
Author names:	Priyanka, Priya Singh
Name of Conference:	9th International Conference on Innovative Computing and Communication (ICICC-2026) – A Flagship Conference
Conference Dates with Venue:	Shaheed Sukhdev College of Business Studies, Delhi University, February 6–7, 2026
Conference Registration:	1055
Paper Status:	Accepted and Presented
Date of Communication:	15 November 2025
Date of Acceptance:	4 December 2025
Date of Publication:	Awaited

Paper 2:

Title of the Paper: A Comparative Analysis of Machine Learning and Deep Learning Approaches for Credit Card Fraud Detection under Class Imbalance: A Review

Author names: Priyanka, Priya Singh

Name of Conference: International Conference on Data-Driven Insights for National Security: Exploring AI and Statistical Frontiers in India's Socio-Cultural Landscape (DINS-2026)

Conference Dates with Venue: Noida International University (NIU), March 30–31, 2026

Conference Registration: 60

Paper Status: Accepted and Presented

Date of Communication: 05 March 2026

Date of Acceptance: 28 March 2026

Date of Publication: Awaiting

Student Name: Priyanka **Roll No.:** 24/SWE/19

Signature of Student

SUPERVISOR CERTIFICATE

To the best of my knowledge, the above work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere. I further certify that the publication and indexing information given by the student is correct.

Place: Delhi

Ms. Priya Singh

Date: _____

Assistant Professor, Dept. of Software Engineering

NOTE: PLEASE ENCLOSE RESEARCH PAPER ACCEPTANCE/PUBLICATION/COMMUNICATION PROOF ALONG WITH SCOPUS INDEXING PROOF.

PUBLICATION DETAILS AND PROOFS

5/8/26, 11:35 PM
Delhi Technological University Mail - ICICC 2026: Paper Notification for Paper ID 1055



DTU
Delhi Technological
University

24/SWE/19 PRIYANKA <priyanka_24swe19@dtu.ac.in>

ICICC 2026: Paper Notification for Paper ID 1055
 1 message

Thu, Dec 4, 2025 at 8:41 AM

ICICC 2026 <icicc.ui@gmail.com>
 To: Priyanka 24swe <priyanka_24swe19@dtu.ac.in>

9th International Conference on Innovative Computing and Communication (ICICC-2026) – A Flagship Conference

Dear Author(s),

Greetings from - ICICC 2026!

We congratulate you that your paper with submission ID '1055' and Paper Title '**A Hybrid Learning Strategy for Credit Card Fraud Detection Using Calibrated Thresholds**' has been accepted for publication in the **Proceedings of ICICC 2026- Springer LNNS Series** [Indexing: zbMATH, Scopus, DBLP, EI Compendex & SCImago - Approved]. This acceptance means your paper is among the top 15% of the papers received/reviewed.

To secure your participation in this prestigious event, we request that you complete your early-bird registration by 10th December 2025 (as mentioned on the website, 05th December 2025). Few slots left. Kindly do the registration on time.

You are requested to do the registration as soon as possible and submit the following documents to icicc.ui@gmail.com at the earliest.

1. Final Camera-Ready Copy (CRC) as per the springer format- (See <https://icicc-conf.com/downloads>)
2. Copy of e-receipt of registration fees. (For Registration, see <https://icicc-conf.com/registrations>)
3. The final revised copy of your paper should also be uploaded via Microsoft CMT.
4. Do not create another ID, just upload the revised version on the existing Microsoft CMT ID only

The reviewers' comments are given at the bottom of this letter; please improve your paper as per the reviewers comments. **While preparing the final CRC manuscript, kindly check the following link under the download section tab named manuscript guidelines:**

<https://icicc-conf.com/registrations>

and it is suggested to cite the relevant latest papers matching the area of your current research paper.

First Conference Acceptance Proof



Money Sent Successfully

₹12,000 

Rupees Twelve Thousand Only

To: Ui Education

UPI ID: 9891296774@kotak

UE

From: Priyanka

ICICI Bank - 1625

8

UPI Ref No: 296482759499

12:26 PM, 10 Dec 2025

paytm | Powered by  |  YES BANK

Fee Receipt for the paper titled, A Hybrid Learning Strategy for Credit Card Fraud Detection Using Calibrated Thresholds



Certificate of Presentation for the paper titled, A Hybrid Learning Strategy for Credit Card Fraud Detection Using Calibrated Thresholds



The conference will be conducted in a hybrid mode, and all paper submissions must be made through the Microsoft CMT portal

Submit Paper

6034 Citations

25 H5 - Index

31 H - Index

2616 No. of Published Papers in Proceedings

79 No. of Keynotes

15 No. of Workshops

20% Acceptance Rate

50 No. of Participating Countries

34 No. of Volumes Published

125 No. of Published Papers in Extended Journals (Scopus)

75 No. of Published Papers in Extended Journals (SCIE)

5300 Total No. of Participation

ICICC Conference H-Index

About Conference

Downloads

International Conference on Innovative Computing and Communication (ICICC-2026) is organized to bring together innovative scientists, professors, research scholars, students, and industrial experts in the field of Computing and Communication to a common forum. The primary goal of the conference is to promote the exchange of innovative scientific information between researchers, developers, engineers, students, and practitioners. Another goal is to promote the transformation of fundamental research into institutional and industrialized research and to convert applied exploration into the real-time application. Overall, the conference will provide the researchers and attendees with prospects for national and international collaboration and networking among universities and institutions from India and abroad for promoting research. ICICC-2026 will be held at Shaheed Sukhdev College of Business Studies, University of Delhi, New Delhi. All the accepted papers (after double-blinded peer review) will be published as proceedings of ICICC-2026 in Springer LNNS Series (Approved)

[Paper Submission](#)

9th International Conference on Innovative Computing and Communication (ICICC-2026) – A Flagship Conference
(Scopus Indexed Proof)

49

5/8/26, 11:36 PM

Delhi Technological University Mail - PID: 60 - Final Acceptance for Oral Presentation at DINS-2026



24/SWE/19 PRIYANKA <priyanka_24swe19@dtu.ac.in>

PID: 60 - Final Acceptance for Oral Presentation at DINS-2026

1 message

Microsoft CMT <noreply@msr-cmt.org>
 To: "Priyanka ." <priyanka_24swe19@dtu.ac.in>
 Cc: dins2026@niu.edu.in

Sat, Mar 28, 2026 at 2:05 PM

Dear Author(s),

Greetings!

We are pleased to inform you that your manuscript "A Comparative Analysis of Machine Learning and Deep Learning Approaches for Credit Card Fraud Detection under Class Imbalance" has been finally accepted for oral presentation at the International Conference on "Data-Driven Insights for National Security: Exploring AI and Statistical Frontiers in India's Socio-Cultural Landscape, DINS-2026" on 30th - 31st March 2026.

We would be delighted if you could join us in person to present your paper, engage with our team, and be part of the conference. You will shortly receive the program schedule and meeting links from our team.

Please note that if any formatting corrections or content changes are required for publication, we will notify you via dins2026@niu.edu.in.

Thank you for choosing DINS-2026 to share your research. We look forward to your participation!

Warm Regards,
 Organizing Committee
 DINS-2026
 Noida International University

Please do not reply to this email as it was generated from an email account that is not monitored.

To stop receiving conference emails, you can check the 'Do not send me conference email' box from your User Profile.

Microsoft respects your privacy. To learn more, please read our [Privacy Statement](#).

Microsoft Corporation
 One Microsoft Way
 Redmond, WA 98052

Second Conference Acceptance Proof

 Noida International University Plot 1, Sector-17 A, Yamuna Expressway, Gautam Budh Nagar, Gautam Buddh Nagar	
Receipt	
Admission No. :	Receipt No. : 1892
Student's Name : Priyanka	Receipt Date : 21-04-2026
Father's Name :	Sem/Year : -
Course/Branch : -	Batch : 2026-2027
Head	Amount (Rs.)
INTERNATIONAL CONFERENCE FEE	7,999.00
Receipt Rs. : Rupees Seven Thousand Nine Hundred Ninety-Nine Only	Total : 7,999.00
Payment By : Online Pay	
Cheque/Order ID : TID : 399551318173, Transaction No. : 399551318173, Card Type :	
Remark : 399551318173 International Conference (DINS -2026) Registration	
*Cheques are subject to realisation *Note : Duplicate copy of Receipt can be obtained against a payment of Rs. 100 *Note : Fee/Amount once deposits will neither be refunded nor adjusted under any circumstances. In case of cheque being dishonored a sum of Rs.1000/- will be charged towards the bank processing charges	
HKRAM CHAUDHAN (Account Officer)	

Fee Receipt for the paper titled, A Comparative Analysis of Machine Learning and Deep Learning Approaches for Credit Card Fraud Detection under Class Imbalance: A Review



Certificate of Presentation for the paper titled, A Comparative Analysis of Machine Learning and Deep Learning Approaches for Credit Card Fraud Detection under Class Imbalance: A Review



Click here to download Template for Manuscript Submission to DINS-2026

Paper Submission Link (CMT Portal)

<https://cmt3.research.microsoft.com/DINS2026>

CMT Acknowledgment

The Microsoft CMT service was used for managing the peer-reviewing process for this conference. This service was provided for free by Microsoft and they bore all expenses, including costs for Azure cloud services as well as for software development and support.

Important Dates

Full Paper Submission	05th March 2026 (Hard Deadline)
Notification of Acceptance	15th March 2026
Submission of the revised paper through the CMT Portal:	20th March 2026
Registration Deadline	25th March 2026
Conference Dates	30th - 31st March 2026

All presented papers will be published in Scopus-indexed platform

International Conference on Data-Driven Insights for National Security: Exploring AI and Statistical Frontiers in India's Socio-Cultural Landscape, DINS-2026
(Scopus Indexed Proof)

**DELHI TECHNOLOGICAL UNIVERSITY**

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-110042

PLAGIARISM VERIFICATION

Title of the Thesis	A Comparative and Hybrid Ensemble Framework for Credit Card Fraud Detection Using Machine Learning and Deep Learning Techniques
Total Pages	_____
Name of the Scholar	Priyanka
Supervisor(s)	(1) Ms. Priya Singh
Department	Software Engineering

This is to report that the above thesis was scanned for similarity detection. Process and outcome is given below:

Software Used	_____
Similarity Index	_____
Total Word Count	_____
Date	_____

Candidate's Signature

Signature of Supervisor(s)

PLAGIARISM REPORT



Page 1 of 51 - Cover Page

SubmissionID trnoid::27535:142398941

Priyanka R

mtech_thesis_priyanka

Mtech Thesis

Document Details

Submission ID

trnoid::27535:142398941

Submission Date

Jun 10, 2026, 10:53 AM GMT+5:30

Download Date

Jun 10, 2026, 11:00 AM GMT+5:30

File Name

mtech_thesis_priyanka.pdf

File Size

523.6 KB

43 Pages

11,968 Words

66,322 Characters



Page 1 of 51 - Cover Page

SubmissionID trnoid::27535:142398941

8% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Submitted works

Match Groups

-  **89 Not Cited or Quoted 7%**
Matches with neither in-text citation nor quotation marks
-  **0 Missing Quotations 0%**
Matches that are still very similar to source material
-  **6 Missing Citation 1%**
Matches that have quotation marks, but no in-text citation
-  **0 Cited and Quoted 0%**
Matches with in-text citation present, but no quotation marks

Top Sources

- 5%  Internet sources
- 6%  Publications
- 0%  Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Priyanka R

mtech_thesis_priyanka

 Mtech Thesis

Document Details

Submission ID

trnoid::27535:142398941

Submission Date

Jun 10, 2026, 10:53 AM GMT+5:30

Download Date

Jun 10, 2026, 11:00 AM GMT+5:30

File Name

mtech_thesis_priyanka.pdf

File Size

523.6 KB

43 Pages

11,968 Words

66,322 Characters

0% detected as AI

The percentage indicates the combined amount of likely AI-generated text as well as likely AI-generated text that was also likely AI-paraphrased.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.



0 AI-generated only 0%

Likely AI-generated text from a large-language model.



0 AI-generated text that was AI-paraphrased 0%

Likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

Disclaimer

Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (i.e., our AI models may produce either false positive results or false negative results), so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

AUTHOR'S BIOGRAPHY

Priyanka is currently pursuing a Master of Technology (M.Tech) in Software Engineering from Delhi Technological University (DTU), New Delhi. She completed her Bachelor's degree in Computer Science and Engineering, where she developed a strong foundation in software development, machine learning, data analytics, and intelligent systems.



Her research interests include machine learning, deep learning, anomaly detection, fraud analytics, and artificial intelligence applications in financial systems. During her postgraduate studies, she has worked extensively on credit card fraud detection, focusing on class-imbalanced data, ensemble learning, anomaly detection techniques, and hybrid machine learning frameworks for improving fraud detection performance.

She has contributed to research in the domain of financial fraud detection, including studies on comparative analysis of machine learning and deep learning approaches and the development of hybrid ensemble models for fraud detection. Her work emphasizes enhancing detection accuracy, reducing false positives, improving model robustness, and developing practical AI-driven solutions for secure financial transaction systems.